

ADAPTIVE CROSS APPROXIMATION FOR ILL-POSED PROBLEMS*

T. MACH[†], L. REICHEL[‡], M. VAN BAREL[§], AND R. VANDEBRIL[¶]

Abstract. Integral equations of the first kind with a smooth kernel and perturbed right-hand side, which represents available contaminated data, arise in many applications. Discretization gives rise to linear systems of equations with a matrix whose singular values cluster at the origin. The solution of these systems of equations requires regularization, which has the effect that components in the computed solution connected to singular vectors associated with small singular values are damped or ignored. In order to compute a useful approximate solution typically approximations of only a fairly small number of the largest singular values and associated singular vectors of the matrix are required. The present paper explores the possibility of determining these approximate singular values and vectors by adaptive cross approximation. This approach is particularly useful when a fine discretization of the integral equation is required and the resulting linear system of equations is of large dimensions, because adaptive cross approximation makes it possible to compute only fairly few of the matrix entries.

Key words. ill-posed problem, inverse problem, sparse discretization, regularization, adaptive cross approximation

1. Introduction. This paper considers the approximate solution of Fredholm integral equations of the first kind,

$$\int_{\Omega_1} \kappa(s, t)x(t)dt = g(s), \quad s \in \Omega_2, \quad (1.1)$$

with a smooth kernel κ . The Ω_i are subsets of \mathbb{R}^{d_i} for some positive integers d_i , $i = 1, 2$. Integral equations of this form arise in many applications, including remote sensing, computerized tomography, and image restoration. The solution of (1.1) is a so-called ill-posed problem. A reason for this is that the singular values of the integral operator cluster at the origin; see, e.g., [10, 18].

Discretization of (1.1) by a Galerkin, Petrov–Galerkin, or Nyström method yield a linear system of equations

$$A\mathbf{x} = \mathbf{g}, \quad A \in \mathbb{R}^{n \times n}, \quad \mathbf{g} \in \mathbb{R}^n, \quad (1.2)$$

with a matrix with many singular values close to the origin. The matrix may be very ill-conditioned already for small to moderate values of n , where we measure the

* The research was partially supported by the Research Council KU Leuven, project OT/10/038 (Multi-parameter model order reduction and its applications), PF/10/002 Optimization in Engineering Centre (OPTEC), CREA-13-012 (Can Unconventional Eigenvalue Algorithms Supersede the State of the Art), OT/11/055 (Spectral Properties of Perturbed Normal Matrices and their Applications), by the Fund for Scientific Research–Flanders (Belgium), G.0828.14N (Multivariate polynomial and rational interpolation and approximation), G.0342.12N (Reestablishing Smoothness for Matrix Manifold Optimization via Resolution of Singularities), and by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office, Belgian Network DYSCO (Dynamical Systems, Control, and Optimization). This research also is supported in part by NSF grant DMS-1115385.

[†]Department of Computer Science, KU Leuven, Celestijnenlaan 200A, 3001 Leuven (Heverlee), Belgium. E-mail: thomas.mach@cs.kuleuven.be.

[‡]Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA. E-mail: reichel@math.kent.edu.

[§]Department of Computer Science, KU Leuven, Celestijnenlaan 200A, 3001 Leuven (Heverlee), Belgium. E-mail: marc.vanbarel@cs.kuleuven.be.

[¶]Department of Computer Science, KU Leuven, Celestijnenlaan 200A, 3001 Leuven (Heverlee), Belgium. E-mail: raf.vandebril@cs.kuleuven.be.

conditioning as the ratio between the largest and smallest singular values of A . In fact, A may be singular. Linear systems of equations with a matrix of this kind are commonly referred to as discrete ill-posed problems. We will for notational simplicity in this paper assume the matrix A to be square, however, the methods described can also be applied after minor modifications when A is rectangular, in which case the linear system of equations (1.2) may be replaced by a least-squares problem.

In many applications, the right-hand side vector \mathbf{g} represents measured data and is contaminated by measurement and discretization errors. Due to these errors and the ill-conditioning of A , straightforward solution of (1.2) typically yields a computed solution that is severely contaminated by propagated error and is therefore not useful. To circumvent this difficulty, the system (1.2) is commonly replaced by a nearby problem that is less sensitive to the error in \mathbf{g} . This replacement is referred to as regularization. The possibly most popular regularization methods include truncated singular value decomposition (TSVD) and Tikhonov regularization.

Define the singular value decomposition (SVD)

$$A = U\Sigma V^T, \quad (1.3)$$

where $U = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n] \in \mathbb{R}^{n \times n}$ and $V = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n] \in \mathbb{R}^{n \times n}$ are orthogonal matrices and

$$\Sigma = \text{diag}[\sigma_1, \sigma_2, \dots, \sigma_n] \in \mathbb{R}^{n \times n}$$

is a diagonal matrix. Its nontrivial entries are the singular values of A ; they are ordered according to $\sigma_1 \geq \sigma_2 \geq \dots \sigma_n \geq 0$. The columns of U and V are commonly referred to as the left and right singular vectors of A , respectively. The superscript T denotes transposition. The matrix

$$A_k = \sum_{j=1}^k \sigma_j \mathbf{u}_j \mathbf{v}_j^T \quad (1.4)$$

is a closest matrix of rank at most k to A in the spectral norm; see, e.g., [15]. The TSVD method determines, for some suitable $k \geq 0$, the solution of minimal Euclidean norm, denoted by \mathbf{x}_k , of the least-squares problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|A_k \mathbf{x} - \mathbf{g}\|. \quad (1.5)$$

Here and throughout this paper $\|\cdot\|$ stands for the Euclidean vector norm or the spectral matrix norm. The parameter k is a regularization parameter that determines how many singular values and vectors of A are used to compute the approximate solution \mathbf{x}_k of (1.2).

Tikhonov regularization replaces the system (1.2) by the penalized least-squares problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \{\|A\mathbf{x} - \mathbf{g}\|^2 + \mu\|\mathbf{x}\|^2\}, \quad (1.6)$$

which has a unique solution \mathbf{x}_μ for any positive value of the regularization parameter μ . Substituting the SVD (1.3) into (1.6) shows that Tikhonov regularization dampens the contributions to \mathbf{x}_μ of singular values and vectors with large index k the most; increasing $\mu > 0$ results in more damping. We refer to [8, 10, 14, 18, 19, 23, 26] for details and computed examples with these regularization methods.

The determination of suitable values of the regularization parameters, k in (1.5) and μ in (1.6), is important for the quality of the computed approximate solution. Several methods have been described in the literature including the discrepancy principle, the L-curve criterion, and generalized cross validation; see [6, 25, 27] for recent discussions of their properties and illustrations of their performance. Regularization methods typically require that regularized solutions for several parameter values be computed and compared in order to determine a suitable value.

The present paper is concerned with the situation when the data vector \mathbf{g} in (1.2) is of high dimension. Then the matrix A is large. The repeated solution of (1.6) can be carried out by iterative methods; see, e.g., [8, 10, 14, 19, 23]. These methods require the evaluation of matrix-vector products with A , and possibly with A^T as well, and this can be expensive when A is large. Moreover, all entries of the matrix have to be computed. There are iterative methods for computing the first k singular values and associated singular vectors of the matrix A ; see, e.g., [2, 22, 24]. These methods also require matrix-vector product evaluations with A and A^T , as well as the evaluation of all matrix elements.

Cross approximation, sometimes also referred to as skeleton approximation, of matrices has been proposed as an approach to approximate a large dense matrix by a matrix of low rank; see, e.g., [4, 5, 13, 16, 29] and references therein. This method seeks to select a subset of k rows and columns of the matrix A to obtain a matrix $M_k \in \mathbb{R}^{n \times n}$ of rank at most k so that $\|A - M_k\|$ is small. Due to the optimality of the SVD of A , we have

$$\|A - A_k\| \leq \|A - M_k\|.$$

However, M_k is much cheaper to compute than A_k . In particular, the determination of M_k does not require that all entries of A be evaluated. The good performance of cross approximation for the approximation of a large matrix A by a matrix M_k of low rank k is well documented in the literature; see, e.g., [4, 5, 13, 16, 29]. However, we are not aware of applications of cross approximation to the solution of discretized Fredholm integral equations of the first kind (1.1) with a smooth kernel. It is the purpose of the present paper to discuss this application.

This paper is organized as follows. Section 2 discusses an algorithm for adaptive cross approximation. Its use in the TSVD method for computing an approximate solution of (1.2) is described in Section 3 and its application in the context of Tikhonov regularization is considered in Section 4. We discuss the TSVD and Tikhonov regularization methods when the regularization parameters k and μ , respectively, are determined by the discrepancy principle. However, other approaches to determine these parameters can also be applied; see, e.g., [6, 10, 19, 25, 27] for discussions and comparisons of a variety of methods for determining the regularization parameters. Section 5 presents computed examples and concluding remarks can be found in Section 6.

2. Adaptive cross approximation. The aim of cross or skeleton approximation of a large matrix A is to determine a matrix M_k of (low) rank at most k , such that M_k approximates A sufficiently well and can be computed much more efficiently than the matrix (1.4). This goal is reached by carefully selecting k rows and columns of the matrix A , called skeletons. The cross approximation is called adaptive when the choice and number of skeletons is determined by properties of A during the computations; see [3, 5]. This section considers three versions of adaptive cross approximation to

reduce square nonsymmetric matrices, symmetric indefinite matrices, and symmetric positive (semi)definite matrices.

Let us first consider the case when the matrix $A \in \mathbb{R}^{n \times n}$ is nonsymmetric. We select k rows of A with row indices $\mathbf{i} \subset \{1, \dots, n\}$ and denote the submatrix so obtained by $A_{(\mathbf{i},:)}$. Similarly, we select k columns of A with column indices $\mathbf{j} \subset \{1, \dots, n\}$ and denote the submatrix so defined by $A_{(:,\mathbf{j})}$. These k rows and columns have a common “core matrix” $A_{(\mathbf{i},\mathbf{j})}$. Assuming that this core matrix is nonsingular, the rows \mathbf{i} and columns \mathbf{j} of the matrix

$$M_k = A_{(\mathbf{i},:)} A_{(\mathbf{i},\mathbf{j})}^{-1} A_{(:,\mathbf{j})}$$

equal the corresponding rows and columns of A . When A is of rank k , we have $M_k = A$.

Assume that the matrix A can be approximated well by a matrix of rank k and that the row and column indices \mathbf{i} and \mathbf{j} are chosen in a suitable way. Then we can expect M_k to be an accurate approximation of A . It is shown by Goreinov et al. [16] that if A can be approximated sufficiently well by a matrix of low rank, then under suitable conditions a cross approximation M_k that approximates A well exists. However, it is not described how to choose suitable index sets \mathbf{i} and \mathbf{j} . Goreinov et al. [17] explain that the index sets \mathbf{i} and \mathbf{j} should be chosen so that $A_{(\mathbf{i},\mathbf{j})}$ is a submatrix of A of maximal volume; i.e., the modulus of the determinant is maximal. However, the determination of such a submatrix is a difficult problem. We therefore seek to determine a low-rank matrix M_k that is a sufficiently accurate approximation of A in a different manner. All existing methods for constructing such a matrix M_k use a greedy approach to successively compute rank-one approximations or skeletons from a fairly small number of skeletons. Suppose that we already have computed an approximation M_{k-1} of rank at most $k-1$ of A , where M_{k-1} is a sum of $k-1$ skeletons, i.e.,

$$M_{k-1} = \sum_{l=1}^{k-1} \mathbf{w}_l^{(c)} (\mathbf{w}_l^{(r)})^T.$$

To compute the next skeleton, a row index i^* and a column index j^* have to be determined. This is done by looking for the index of the maximum element in magnitude (pivot) in the previously computed vectors $\mathbf{w}_{k-1}^{(c)}$ (for pivot index i^*) and $\mathbf{w}_{k-1}^{(r)}$ (for pivot index j^*). In the beginning an arbitrary row of A can be chosen. We will choose the first row in the computed examples of Section 5. In the simplest form of cross approximation, the computation of the vectors $\mathbf{w}_k^{(c)}$ and $\mathbf{w}_k^{(r)}$ only requires the entries in row i^* and column j^* of A and the entries of already computed skeletons:

$$\begin{aligned} (\mathbf{w}_k^{(r)})_j &= A_{i^*,j} - \sum_{l=1}^{k-1} (\mathbf{w}_l^{(c)})_{i^*} (\mathbf{w}_l^{(r)})_j, \quad \delta = (\mathbf{w}_k^{(r)})_{j_k^*}, \\ (\mathbf{w}_k^{(c)})_i &= \frac{1}{\delta} \left(A_{i,j_k^*} - \sum_{l=1}^{k-1} (\mathbf{w}_l^{(c)})_i (\mathbf{w}_l^{(r)})_{j_k^*} \right). \end{aligned}$$

A new skeleton is obtained from the remainder,

$$R_k = A - \sum_{l=1}^k \mathbf{w}_l^{(c)} (\mathbf{w}_l^{(r)})^T,$$

without explicitly computing R_k .

In general, the number of skeletons required, k , is not known in advance. Therefore a stopping criterion is introduced to adaptively determine k . If the stopping criterion would be based only on the value of the pivot element that is chosen in each step, then this criterion would fail when the matrix A is block diagonal,

$$A = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix},$$

with $A_{11} \in \mathbb{R}^{n_1 \times n_1}$, $A_{22} \in \mathbb{R}^{n_2 \times n_2}$, and $n = n_1 + n_2$. Let the initial row and column indices i^* and j^* , respectively, belong to one of the index sets $I_1 = \{1, \dots, n_1\}$ or $I_2 = \{n_1 + 1, \dots, n_1 + n_2\}$. Then the next chosen pivot indices will be elements of the same index set. Therefore, at most one of the two diagonal blocks, A_{11} or A_{22} , will be approximated.

Most stopping criteria described in the literature, see, e.g., [3, 4, 5], only use the computed rows and columns and therefore cannot avoid this difficulty. A new stopping criterion was developed in [13], which also includes t randomly chosen matrix entries A_{i_l, j_l} , for $l = 1, \dots, t$, with $i_l, j_l \in \{1, 2, \dots, n\}$. When a new skeleton is determined, the values of these entries are updated by subtraction of the available skeletons,

$$(R_k)_{i_l, j_l} = (R_{k-1})_{i_l, j_l} - (\mathbf{w}_k^{(c)})_{i_l} (\mathbf{w}_k^{(r)})_{j_l}, \quad (2.1)$$

with $(R_0)_{i_l, j_l} = A_{i_l, j_l}$. If the following condition holds,

$$|(R_k)_{i_l, j_l}| \leq \tau \quad \forall l = 1, \dots, t, \quad (2.2)$$

for a user-specified parameter-value τ , then the algorithm will stop. The number of entries t considered in this stopping criterion is a percentage of the total number of entries. The choice of this percentage should depend on properties of the matrix A ; see [13] for further details.

Algorithm 1 below is an adaptation and reformulation of the algorithm described in [13]. It is applicable to nonsymmetric matrices $A \in \mathbb{R}^{n \times n}$. In the algorithm, the pivot index i_k^* is determined by the element of largest magnitude of $\mathbf{w}_k^{(c)}$ or by the magnitudes of $(R_{k-1})_{i_l, j_l}$, $l = 1, 2, \dots, t$. A variant could be to always determine i_k^* by using the magnitude of the elements $(R_{k-1})_{i_l, j_l}$ only.

When the matrix A is symmetric, we would like the approximant M_k also to be symmetric. Let us first consider the case when A is symmetric and indefinite. Algorithm 1 can be adapted as follows to compute an approximant M_k of the form

$$M_k = \sum_{l=1}^k \delta_l^{-1} \mathbf{w}_l \mathbf{w}_l^T.$$

Assume that the stopping criterion is not satisfied. When $j_k^* = i_k^*$ only one skeleton $\delta_k^{-1} \mathbf{w}_k \mathbf{w}_k^T$ is added to M_{k-1} , while when $j_k^* \neq i_k^*$ two skeletons are added at the same time. They are determined by the two rows and columns with indices i_k^* and j_k^* . The set Z_i (see Algorithm 1) is then updated according to

$$Z_i \cup \{i_p^*, j_p^*\}.$$

The next possible value for i_{k+1}^* is taken as the maximum absolute value of the two columns

$$(R_{k-1})_{(i, i_k^*)} \text{ and } (R_{k-1})_{(i, j_k^*)} \quad \text{with } i \in I \setminus Z_i.$$

Algorithm 1 Adaptive Cross Approximation (ACA)

Choose $i_1^* \in I := \{1, \dots, n\}$, tolerance τ , number of points t in the stopping criterion. Set $Z_i := \{\}$, $k := 1$, $J := \{1, \dots, n\}$, **stopcrit:=false**.

while Not **stopcrit** **do**

 Compute the maximal entry in modulus of the row

$$\begin{aligned} (\mathbf{w}_k^{(r)})_j &:= A_{i_k^*, j} - \sum_{l=1}^{k-1} (\mathbf{w}_l^{(c)})_{i_k^*} (\mathbf{w}_l^{(r)})_j, & j \in J, \\ j_k^* &:= \arg \max_{j \in J} |(\mathbf{w}_k^{(r)})_j|, \\ \delta_k &:= (\mathbf{w}_k^{(r)})_{j_k^*}. \end{aligned}$$

if $|\delta_k| \leq \tau$ **then**

 Set $i_k^* := i_{l^*}$ with $l^* := \arg \max_l |(R_{k-1})_{i_l, j_l}|$.

else

 Set $Z_i := Z_i \cup \{i_k^*\}$.

 Compute the entries of the following vector

$$(\mathbf{w}_k^{(c)})_i := \frac{1}{\delta_k} \left(A_{i, j_k^*} - \sum_{l=1}^{k-1} (\mathbf{w}_l^{(c)})_i (\mathbf{w}_l^{(r)})_{j_k^*} \right), \quad i \in I.$$

 Compute the maximal entry in modulus of the column

$$i_{k+1}^* := \arg \max_{i \in I \setminus Z_i} |(\mathbf{w}_k^{(c)})_i|.$$

stopcrit := $\max_{1 \leq l \leq t} |(R_k)_{i_l, j_l}| \leq \tau$.

 Set $k = k + 1$.

end if

end while

The computed approximant M_k is an incomplete LDL^T-type factorization of the symmetrically pivoted matrix A .

When A is symmetric and positive (semi)definite, the stopping criterion can be simplified. Instead of considering t randomly chosen elements of the matrix A , we only have to look at the diagonal entries. In step k , we determine the largest element in absolute value of $(R_{k-1})_{(i,i)}$, i.e.,

$$i_k^* = j_k^* = \arg \max_i |(R_{k-1})_{(i,i)}|.$$

Note that in this case all the δ_k -values will be positive and the approximant M_k is an incomplete Cholesky factorization of the symmetrically pivoted matrix A . The following result sheds some light onto the modification of Algorithm 1 for symmetric positive semidefinite matrices.

PROPOSITION 2.1. *Let $A \in \mathbb{R}^{n \times n}$ be symmetric positive semidefinite. Each one of the matrices R_1, R_2, \dots, R_k , analogous to (2.1), generated by the above described modification of Algorithm 1 for positive semidefinite matrices, is symmetric positive*

semidefinite.

Proof. We only have to consider one step of the algorithm. Let $i_1^* = 1$. The matrix $C_\gamma = I - \gamma \mathbf{e}_1 \mathbf{e}_1^T$ with $0 < \gamma < 1$ is symmetric positive definite and by Sylvester's law of inertia $C_\gamma A C_\gamma$ is positive semidefinite. A perturbation argument shows that $C_1 A C_1$ also is positive semidefinite. The proposition now follows from the observation that $R_1 = C_1 A C_1$. \square

3. The TSVD method. We first consider the situation when the matrix $A \in \mathbb{R}^{n \times n}$ is nonsymmetric. Application of k steps of Algorithm 1 gives two matrices,

$$W_k^{(c)} = [\mathbf{w}_1^{(c)}, \mathbf{w}_2^{(c)}, \dots, \mathbf{w}_k^{(c)}] \in \mathbb{R}^{n \times k}, \quad W_k^{(r)} = [\mathbf{w}_1^{(r)}, \mathbf{w}_2^{(r)}, \dots, \mathbf{w}_k^{(r)}] \in \mathbb{R}^{n \times k}$$

such that

$$A \approx M_k = W_k^{(c)} (W_k^{(r)})^T. \quad (3.1)$$

Introduce the QR factorizations

$$W_k^{(c)} = Q_k^{(c)} R_k^{(c)}, \quad W_k^{(r)} = Q_k^{(r)} R_k^{(r)}. \quad (3.2)$$

where the matrices $Q_k^{(c)}, Q_k^{(r)} \in \mathbb{R}^{n \times k}$ have orthonormal columns and $R_k^{(c)}, R_k^{(r)} \in \mathbb{R}^{k \times k}$ are upper triangular. The latter matrices may be numerically singular. These QR factorizations can be computed by the Householder-QR method or by methods that run more efficiently on modern computers; see, e.g., [7, 11, 30] for examples.

We will use the singular value decomposition

$$R_k^{(c)} (R_k^{(r)})^T = \tilde{U}_k \tilde{\Sigma}_k \tilde{V}_k^T, \quad (3.3)$$

where the matrices $\tilde{U}_k, \tilde{V}_k \in \mathbb{R}^{k \times k}$ are orthogonal and $\tilde{\Sigma}_k = \text{diag}[\tilde{\sigma}_1, \tilde{\sigma}_2, \dots, \tilde{\sigma}_k] \in \mathbb{R}^{k \times k}$ has the nontrivial entries $\tilde{\sigma}_1 \geq \tilde{\sigma}_2 \geq \dots \geq \tilde{\sigma}_k \geq 0$. The SVD of M_k is given by

$$M_k = Q_k^{(c)} \tilde{U}_k \tilde{\Sigma}_k \tilde{V}_k^T (Q_k^{(r)})^T. \quad (3.4)$$

Since $n \gg k$ in our applications, the dominating computational effort for determining the SVD (3.4) is the computation of the QR factorizations (3.2).

We first describe the discrepancy principle when applied to the solution of (1.2) by using (1.4), and then discuss necessary modifications required when applying the discrepancy principle to the solution of

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|M_k \mathbf{x} - \mathbf{g}\|. \quad (3.5)$$

Let a fairly accurate bound, ε , for the norm of the error in \mathbf{g} be known, i.e.,

$$\|\mathbf{g} - \mathbf{g}_{\text{exact}}\| \leq \varepsilon, \quad (3.6)$$

where $\mathbf{g}_{\text{exact}}$ denotes the unknown error-free vector associated with \mathbf{g} , and assume that the error-free linear system of equations associated with (1.2),

$$A \mathbf{x} = \mathbf{g}_{\text{exact}}, \quad (3.7)$$

is consistent. This condition is required when the discrepancy is applied to determine a suitable regularized approximate solution of (1.2). We refer to the solution of minimal Euclidean norm of (3.7) by $\mathbf{x}_{\text{exact}}$. The discrepancy principle prescribes that

the truncation index k in (1.4) be chosen as small as possible so that the minimal-norm solution \mathbf{x}_k of (1.5) satisfies

$$\|A\mathbf{x}_k - \mathbf{g}\| \leq \eta\varepsilon, \quad (3.8)$$

where $\eta \geq 1$ is a user-supplied constant independent of ε . The truncation index k typically increases as ε decreases to zero. Engl et al. [10] show in a Hilbert space setting that \mathbf{x}_k converges to $\mathbf{x}_{\text{exact}}$ as ε decreases to zero.

Since the matrix A is not available, we instead apply the discrepancy principle to the solution of (3.5). Let $\tilde{\sigma}_p$ be the smallest positive singular value of M_k and define, for $1 \leq \ell \leq p$, the matrices

$$\tilde{\Sigma}_k^{(\ell)} = \text{diag}[\tilde{\sigma}_1, \tilde{\sigma}_2, \dots, \tilde{\sigma}_\ell, 0, \dots, 0] \in \mathbb{R}^{k \times k}.$$

and

$$M_k^{(\ell)} = Q_k^{(c)} \tilde{U}_k \tilde{\Sigma}_k^{(\ell)} \tilde{V}_k^T (Q_k^{(r)})^T. \quad (3.9)$$

Let $\mathbf{x}_k^{(\ell)}$ denote the minimal-norm solution of the least-squares problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|M_k^{(\ell)} \mathbf{x} - \mathbf{g}\|. \quad (3.10)$$

We easily can compute $\mathbf{x}_k^{(\ell)}$ with the aid of the decomposition (3.9) as follows. Note that the minimization problem (3.10) is equivalent to

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|Q_k^{(c)} \tilde{U}_k \tilde{\Sigma}_k^{(\ell)} \tilde{V}_k^T (Q_k^{(r)})^T \mathbf{x} - Q_k^{(c)} (Q_k^{(c)})^T \mathbf{g}\|^2 + \|(I - Q_k^{(c)} (Q_k^{(c)})^T) \mathbf{g}\|^2.$$

Let $\mathbf{y}_k^{(\ell)}$ be the least-squares solution of minimal norm of the reduced problem

$$\min_{\mathbf{y} \in \mathbb{R}^k} \|\tilde{\Sigma}_k^{(\ell)} \mathbf{y} - \tilde{U}_k^T (Q_k^{(c)})^T \mathbf{g}\|.$$

Then $\mathbf{x}_k^{(\ell)} = Q_k^{(r)} \tilde{V}_k \mathbf{y}_k^{(\ell)}$. The discrepancy principle suggests that we choose $0 \leq \ell \leq p$ as small as possible so that

$$\|M_k \mathbf{x}_k^{(\ell)} - \mathbf{g}\| \leq \eta\varepsilon, \quad (3.11)$$

where η and ε are the same as in (3.8). Here we assume that such a value of ℓ exists; see below for further comments on this. The left-hand side can be evaluated inexpensively for several values of $0 \leq \ell \leq p$ according to

$$\|M_k \mathbf{x}_k^{(\ell)} - \mathbf{g}\|^2 = \|\tilde{U}_k \tilde{\Sigma}_k \mathbf{y}_k^{(\ell)} - (Q_k^{(c)})^T \mathbf{g}\|^2 + \|(I - Q_k^{(c)} (Q_k^{(c)})^T) \mathbf{g}\|^2.$$

The following result sheds some light on whether an index $1 \leq \ell \leq p$ exists such that (3.11) can be satisfied.

PROPOSITION 3.1. *Let $\tilde{\sigma}_p$ for some $1 \leq p \leq k$ be the smallest positive singular value of the matrix M_k and let the matrix $\tilde{U}_p \in \mathbb{R}^{k \times p}$ be made up of the first p columns of the matrix \tilde{U}_k in (3.4). Then there is a vector $\mathbf{x} \in \mathbb{R}^n$ such that $\|M_k \mathbf{x} - \mathbf{g}\| \leq \eta\varepsilon$ if and only if*

$$\|P_{\mathcal{N}(Q_k^{(c)} \tilde{U}_p)} \mathbf{g}\| \leq \eta\varepsilon, \quad (3.12)$$

where $P_{\mathcal{N}(Q_k^{(c)} \tilde{U}_p)}$ denotes the orthogonal projector onto the null space of $Q_k^{(c)} \tilde{U}_p$. The minimal-norm solution $\mathbf{x}_k^{(p)}$ of (3.10) (with $\ell = p$) satisfies

$$\|M_k \mathbf{x}_k^{(p)} - \mathbf{g}\| = \|P_{\mathcal{N}(Q_k^{(c)} \tilde{U}_p)} \mathbf{g}\|. \quad (3.13)$$

Moreover, $\|M_k \mathbf{x}_k^{(\ell)} - \mathbf{g}\|$ decreases as ℓ increases, and $0 \leq \ell \leq p$.

Proof. Using the SVD of M_k (3.4), we obtain the relation

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|M_k \mathbf{x} - \mathbf{g}\| = \|P_{\mathcal{N}(Q_k^{(c)} \tilde{U}_p)} \mathbf{g}\|,$$

from which (3.12) follows. The equality (3.13) as well as the fact that $\|M_k \mathbf{x}_k^{(\ell)} - \mathbf{g}\|$ decreases when ℓ increases also can be shown with the aid of (3.4). \square

The singular values of M_k are bounded by the singular values of A , i.e., $\sigma_1 \geq \tilde{\sigma}_1$ and $\tilde{\sigma}_k \geq \sigma_n$; see, e.g., [15]. However, we cannot evaluate $\|A - M_k\|$. Therefore, we choose $k \geq 3$ large enough so that

$$\tilde{\sigma}_\ell \geq 10 \tilde{\sigma}_{k-2}. \quad (3.14)$$

The value of ℓ in (3.14) depends on the amount of error in the data \mathbf{g} , and both k and ℓ depend on the matrix A . Typically, it is not known how large k should be chosen in the beginning of the computations. For many problems of interest $10 \leq k \leq 20$ is sufficiently large. If k is found to be too small to satisfy (3.14), then we increase k and update the QR factorizations (3.2) using the method described in [9].

Finally, we note that in order for the discrepancy principle to be able to determine a suitable truncation index ℓ in (3.11), the unknown error-free data vector $\mathbf{g}_{\text{exact}}$ should be in the range of M_k . The distance between $\mathbf{g}_{\text{exact}}$ and the range of M_k depends on how well M_k approximates A . Recall that we assumed $\mathbf{g}_{\text{exact}}$ to be in the range of A . We have found the discrepancy principle to perform better when used with the projected data vector $Q_k^{(c)}(Q_k^{(c)})^T \mathbf{g}$ than with \mathbf{g} . Thus, we let $0 \leq \ell \leq p$ be as small as possible so that

$$\|M_k \mathbf{x}_k^{(\ell)} - Q_k^{(c)}(Q_k^{(c)})^T \mathbf{g}\| \leq \eta \varepsilon. \quad (3.15)$$

The value of ℓ determined in this manner is generally smaller than the value that would have been computed by using (3.11).

We remark that the linear system of equations (1.2) is regularized by first substituting the matrix A by the matrix M_k of rank at most k , and then by replacing the latter by a truncated singular value decomposition $M_k^{(\ell)}$ of rank at most $\ell \leq k$. The number of cross approximation steps k is chosen so that M_k is an adequate approximation of A ; this is the purpose of the requirement (3.14). The truncation index ℓ in (3.9), determined by (3.15), has the effect of setting the $k - \ell$ smallest singular values of M_k to zero and thereby reducing the propagation of the error in \mathbf{g} into the computed approximation $\mathbf{x}_k^{(\ell)}$ of $\mathbf{x}_{\text{exact}}$.

We conclude this section with some comments on the situation when the matrix A is symmetric. The modification of Algorithm 1 outlined at the end of Section 2 for the situation when A is indefinite yields a matrix $W_k = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k] \in \mathbb{R}^{n \times k}$ and a diagonal matrix $D_k = \text{diag}[\delta_1, \delta_2, \dots, \delta_k] \in \mathbb{R}^{k \times k}$ so that

$$M_k = W_k D_k^{-1} W_k^T$$

approximates A . We compute the QR factorization $W_k = Q_k R_k$, where $Q_k \in \mathbb{R}^{n \times k}$ has orthonormal columns and $R_k \in \mathbb{R}^{k \times k}$ is upper triangular, and the SVD

$$R_k D_k^{-1} R_k^T = \tilde{U}_k \tilde{\Sigma}_k \tilde{V}_k^T. \quad (3.16)$$

The matrices in the right-hand side are of the same form as in (3.3). The computations now proceed as described above. Alternatively, one may compute the spectral factorization of $R_k D_k^{-1} R_k^T$ and regularize by removing the eigenvalues of smallest magnitude and associated eigenvectors from this matrix. When all entries of the diagonal matrix D_k in (3.16) are nonnegative, the singular value decomposition of $R_k D_k^{-1/2}$ should be computed instead of (3.16).

4. Tikhonov regularization. To apply this regularization method, we replace the matrix A in (1.6) by the matrix M_k defined by (3.1),

$$\min_{\mathbf{x} \in \mathbb{R}^n} \{ \|M_k \mathbf{x} - \mathbf{g}\|^2 + \mu \|\mathbf{x}\|^2 \}. \quad (4.1)$$

The discrepancy principle applied to the solution of (3.5) prescribes that the regularization parameter $\mu > 0$ be chosen so that the solution $\mathbf{x}_k^{(\mu)}$ of (4.1) satisfies

$$\|M_k \mathbf{x}_k^{(\mu)} - \mathbf{g}\| = \eta \varepsilon, \quad (4.2)$$

where we use the same notation as in Section 3. We refer to this value of the regularization parameter as μ^* . The following result is concerned with the existence of the value μ^* .

PROPOSITION 4.1. *Let p , σ_p , \tilde{U}_p , and the orthogonal projector $P_{\mathcal{N}(Q_k^{(c)} \tilde{U}_p)}$ be defined as in Proposition 3.1. Assume that $M_k^T \mathbf{g} \neq \mathbf{0}$ and define the function*

$$\varphi(\nu) = \mathbf{g}^T (\nu M_k M_k^T + I)^{-2} \mathbf{g}. \quad (4.3)$$

Then, for $\nu > 0$, the solution $\mathbf{x}_k^{(1/\nu)}$ of (4.1) (with $\mu = 1/\nu$) satisfies

$$\varphi(\nu) = \|M_k \mathbf{x}_k^{(1/\nu)} - \mathbf{g}\|^2. \quad (4.4)$$

Moreover, φ is decreasing and convex for $\nu \geq 0$ and the equation

$$\varphi(\nu) = \tau$$

has a unique solution $0 < \nu < \infty$ for any τ such that $\|P_{\mathcal{N}(Q_k^{(c)} \tilde{U}_p)} \mathbf{g}\|^2 < \tau < \|\mathbf{g}\|^2$.

Proof. The proposition has been shown in [8, Theorem 2.1]. We therefore only provide an outline of the proof. The solution of (4.1) is, for any $\mu > 0$, given by

$$\mathbf{x}_k^{(\mu)} = (M_k^T M_k + \mu I)^{-1} M_k \mathbf{g}.$$

Substituting this expression into (4.2) and using the relation $M_k (M_k^T M_k + \mu I)^{-1} M_k^T = M_k M_k^T (M_k M_k^T + \mu I)^{-1}$ gives (4.4) with $\nu = 1/\mu$. The fact that φ is decreasing and convex follows from (4.3). Substituting the SVD of M_k into (4.3) shows the remaining properties of φ . \square

We use the decomposition (3.4) to reduce the minimization problem (4.1) in a similar fashion as in Section 3. Thus, problem (4.1) is equivalent to

$$\min_{\mathbf{y} \in \mathbb{R}^k} \{ \|\tilde{\Sigma}_k \mathbf{y} - \tilde{U}_k^T (\tilde{Q}_k^{(c)})^T \mathbf{g}\|^2 + \mu \|\mathbf{y}\|^2 \}$$

with solution

$$\mathbf{y}_k^{(\mu)} = (\tilde{\Sigma}_k^2 + \mu I)^{-1} \tilde{\Sigma}_k \tilde{U}_k^T (Q_k^{(c)})^T \mathbf{g}.$$

Then $\mathbf{x}_k^{(\mu)} = Q_k^{(r)} \tilde{V}_k \mathbf{y}_k^{(\mu)}$. Substitution into (4.2) gives

$$\|M_k \mathbf{x}_k^{(\mu)} - \mathbf{g}\|^2 = \|\tilde{\Sigma}_k \mathbf{y}_k^{(\mu)} - \tilde{U}_k^T (\tilde{Q}_k^{(c)})^T \mathbf{g}\|^2 + \|(I - Q_k^{(c)} (Q_k^{(c)})^T) \mathbf{g}\|^2. \quad (4.5)$$

The right-hand side can be evaluated inexpensively for any μ -value of interest. A root-finder, such as Newton's method, can be used to determine a value of μ such that $\mathbf{x}_k^{(\mu)}$ satisfies (4.2) or, equivalently, a value $\nu = 1/\mu$ such that

$$\varphi(\nu) = \eta^2 \varepsilon^2.$$

Similarly as in Section 3, it often is preferable to replace the data vector \mathbf{g} by the projected vector $Q_k^{(c)} (Q_k^{(c)})^T \mathbf{g}$ in (4.1) and (4.2). Thus, we solve

$$\min_{\mathbf{x} \in \mathbb{R}^n} \{\|M_k \mathbf{x} - Q_k^{(c)} (Q_k^{(c)})^T \mathbf{g}\|^2 + \mu \|\mathbf{x}\|^2\} \quad (4.6)$$

for $\mathbf{x}_k^{(\mu)}$ and determine $\mu^* := \mu$ so that

$$\|M_k \mathbf{x}_k^{(\mu^*)} - Q_k^{(c)} (Q_k^{(c)})^T \mathbf{g}\| = \eta \varepsilon. \quad (4.7)$$

Analogously to the condition (3.14) for the TSVD method, we require when applying Tikhonov regularization that k is large enough so that

$$\mu^* \geq 100 \tilde{\sigma}_k^2.$$

5. Computed examples. As a proof of concept we will compute regularized solutions for some ill-posed problems from Hansen's regularization MATLAB toolbox [21]. Adaptive cross approximation is used to approximate matrices A by matrices M_k of low rank. Regularization is achieved by the TSVD method applied to (3.15) and by Tikhonov regularization (4.6). The regularization parameters ℓ and μ , respectively, are determined by the discrepancy principle. Throughout this section, $\alpha e-\beta$ stands for $\alpha \cdot 10^{-\beta}$.

Example 5.1. ("Baart" [1, 21]) This example is a Fredholm integral equation of the first kind (1.1) with $\kappa(s, t) = \exp(s \cos(t))$, $g(s) = 2 \sinh(s)/s$, and solution $x(t) = \sin(t)$, where $\Omega_1 = [0, \pi]$ and $\Omega_2 = [0, \pi/2]$. The code `baart` from [21] uses a Galerkin method with piecewise constant test and trial functions to determine the discretized integral operator $A \in \mathbb{R}^{n \times n}$, a discretized scaled approximation $\mathbf{x}_{\text{exact}} \in \mathbb{R}^n$ of $x(t)$, and the corresponding error-free right-hand side vector $\mathbf{g}_{\text{exact}} \in \mathbb{R}^n$. We add the vector $\mathbf{r} \in \mathbb{R}^n$, which models measurement error, to $\mathbf{g}_{\text{exact}}$ to obtain the error-contaminated right-hand side \mathbf{g} in (1.2). The entries of \mathbf{r} are normally distributed with zero mean and scaled so that $\varepsilon := \|\mathbf{r}\| = 1e-4$. We let $n = 2000$ and choose the constant η in (3.8) to be 1.

We compute the approximation $M_k \approx A$ by adaptive cross approximation as described by Algorithm 1. The number of random index pairs, t , used in the stopping criterion of the algorithm is $n/4$, and we denote this set of index pairs by T .

For truncated SVD regularization, we start with $\ell = 1$ and increase ℓ until the residual norm in (3.15),

$$r_\ell := \|M_k \mathbf{x}_k^{(\ell)} - Q_k^{(c)} (Q_k^{(c)})^T \mathbf{g}\|,$$

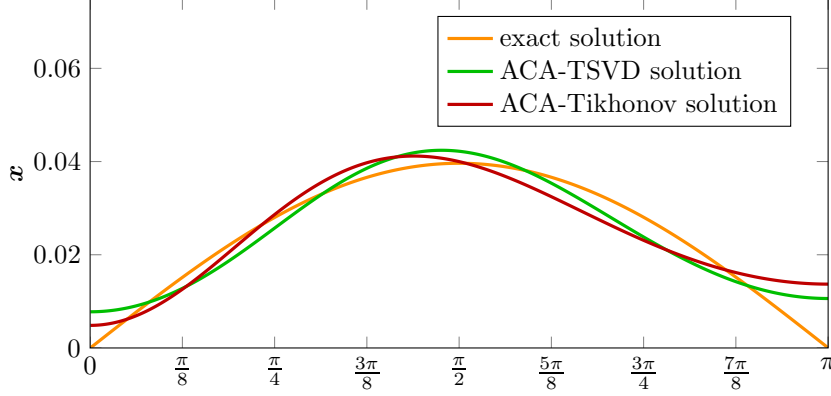


FIG. 5.1. *Example 5.1—“Baart”, $\varepsilon = 1e-4$.*

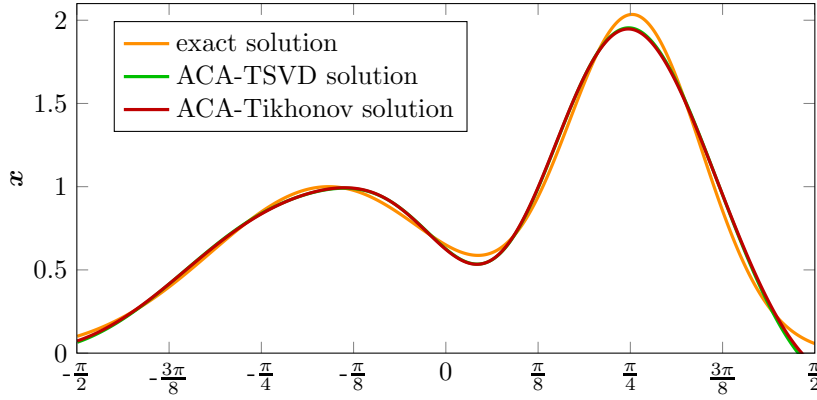


FIG. 5.2. *Example 5.2—“Shaw”, $n = 2000$, $\varepsilon = 1e-4$.*

is smaller than ε .

For Tikhonov regularization, we use the MATLAB function `fminbnd` to find the μ -value for which the solution of (4.6) satisfies fulfills (4.7). Figure 5.1 displays the exact solution, defined by $\mathbf{x}_{\text{exact}}$, and the computed approximate solutions determined by TSVD and Tikhonov regularization. We summarize some key quantities for all examples in Tables 5.1 and 5.2. The vector $\mathbf{x}_{\text{exact}}$ is in the tables denoted by \mathbf{x} .

Example 5.2. (“Shaw” [21, 28]) The kernel in (1.1) is taken to be $\kappa(s, t) = (\cos(s) + \cos(t))(\sin(u)/u)^2$, with $u = \pi(\sin(s) + \sin(t))$, and $\Omega_1 = \Omega_2 = [-\pi/2, \pi/2]$. The solution $x(t)$ is the sum of two Gaussian functions; see [28]. We discretize by a Nyström method on grids with $n = 2000$ and $n = 14500$ grid points. The function `shaw` from [21] determines a matrix $A \in \mathbb{R}^{n \times n}$ and vector $\mathbf{x}_{\text{exact}} \in \mathbb{R}^n$. The error-free right-hand side is defined by $\mathbf{g}_{\text{exact}} = A\mathbf{x}_{\text{exact}}$. Adding Gaussian noise of absolute norm $1e-4$ gives the error-contaminated vector \mathbf{g} .

The computations with Algorithm 1 are carried out similarly as for Example 5.1. The exact and computed regularized solutions for $n = 2000$ are displayed in Figure 5.2. The corresponding results for $n = 14500$ are shown in Figure 5.3. The latter figure and Tables 5.1 and 5.2 illustrate that the method of this paper performs well also for very large matrices.

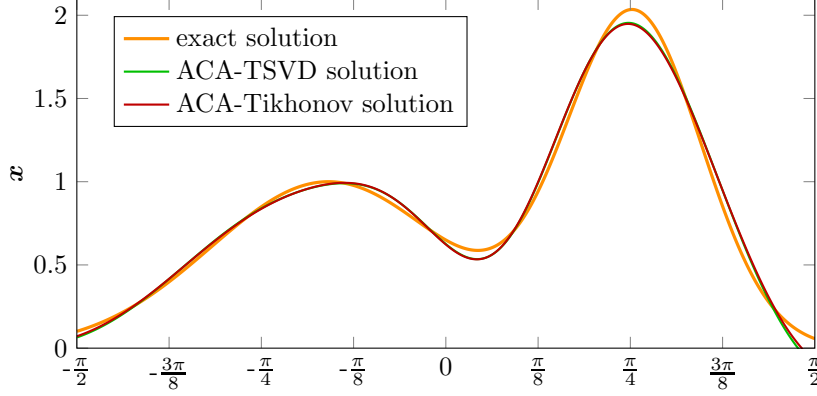


FIG. 5.3. Example 5.2-“Shaw”, $n = 14\,500$, $\varepsilon = 1\,e-4$.

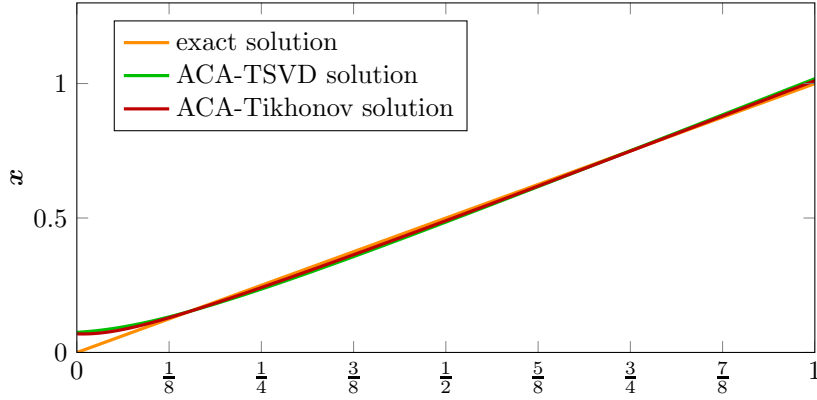


FIG. 5.4. Example 5.3-“Foxgood”, $\varepsilon = 1\,e-2$.

Example 5.3. (“Foxgood” [12, 21]) The integral equation (1.1) is defined by $\kappa(s, t) = \sqrt{s^2 + t^2}$, $\Omega_1 = \Omega_2 = [0, 1]$, and $g(s) = \frac{1}{3}(1 + s^2)^{\frac{3}{2}} - s^3$. It has the solution $x(t) = t$. This example was first used by Fox and Goodwin [12]. The function foxgood from [21] discretizes the integral equation by a Nyström method and gives the matrix $A \in \mathbb{R}^{n \times n}$ and vectors $\mathbf{x}_{\text{exact}}, \mathbf{g}_{\text{exact}} \in \mathbb{R}^n$ similarly as in the previous examples. Gaussian noise $\mathbf{r} \in \mathbb{R}^n$ is added to $\mathbf{g}_{\text{exact}}$ to yield \mathbf{g} . The vector \mathbf{r} is scaled to correspond to an absolute error $1\,e-2$. The results are shown in Figure 5.4.

Example 5.4. (“Gravity” [20, 21]) The integral operator (1.1) is defined by $\kappa(s, t) = d(d^2 + (s - t)^2)^{-\frac{3}{2}}$ with $d = 0.25$ and $\Omega_1 = \Omega_2 = [0, 1]$. The solution $x(t) = \sin(\pi t) + 0.5 \sin(2\pi t)$ is given and defines the right-hand side function g . The matrix $A \in \mathbb{R}^{n \times n}$ and vectors $\mathbf{x}_{\text{exact}}, \mathbf{g}_{\text{exact}} \in \mathbb{R}^n$ are determined by discretizing (1.1) using the function gravity from [21]. We let $n = 2000$ and determine the error-contaminated right-hand side vector $\mathbf{g} \in \mathbb{R}^n$ by adding Gaussian noise that corresponds to an absolute error $1\,e-2$. The results are shown in Figure 5.5.

The matrices in Examples 5.2, 5.3, and 5.4 are symmetric. We also applied the version of adaptive cross approximation for symmetric indefinite matrices described at the end of Section 2 to these matrices. There was no significant difference in the runtime for the symmetric and nonsymmetric adaptive cross approximation algorithms,

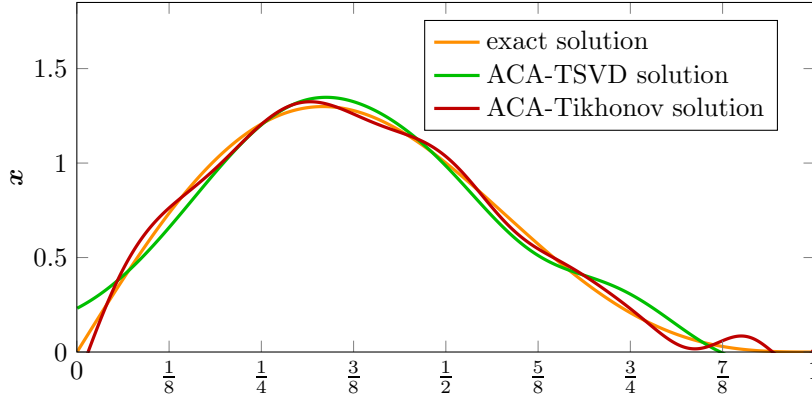


FIG. 5.5. Example 5.4–“Gravity”, $\varepsilon = 1e-2$.

TABLE 5.1
Errors in regularized solution with ACA and TSVD regularization.

Example	n	ε	k	$\ A - M_k\ $	$\ \mathbf{x}_k^{(\ell)} - \mathbf{x}\ /\ \mathbf{x}\ $	ℓ
Ex. 5.1	2 000	1 e−04	7	1.12 e−07	1.15 e−01	4
Ex. 5.2	2 000	1 e−04	11	1.27 e−05	4.76 e−02	12
Ex. 5.2	14 500	1 e−04	11	1.28 e−05	4.76 e−02	7
Ex. 5.2s	2 000	1 e−04	11	2.26 e−05	4.76 e−02	7
Ex. 5.3	2 000	1 e−02	5	4.14 e−04	3.17 e−02	6
Ex. 5.3s	2 000	1 e−02	4	1.22 e−02	4.49 e−02	2
Ex. 5.4	2 000	1 e−02	9	4.11 e−01	8.02 e−02	14
Ex. 5.4s	2 000	1 e−02	11	7.92 e−01	8.18 e−02	4

but the former, of course, requires only about half as much computer memory than the latter. Moreover, the approximate solutions determined by Tikhonov regularization often have the same accuracy when symmetry is exploited, see Table 5.1 and 5.2, where we marked the lines in which we used symmetric cross approximation by ‘s’.

Finally, Tables 5.3 and 5.4 show computations analogous to those reported in Tables 5.1 and 5.2 but without application of adaptive cross approximation. Table 5.3 applies for $n = 2000$ the TSVD method to solve the discrete ill-posed problems of Examples 5.1–5.4. The truncation index k differs from that in Table 5.1, but the quality of the computed approximations of $\mathbf{x}_{\text{exact}}$ is on average about the same as in Table 5.1. Table 5.4 shows results obtained with Tikhonov regularization without adaptive cross approximation for the same problems. The regularization parameter values in Tables 5.4 and 5.2 are very close, and the quality of the computed approximations of $\mathbf{x}_{\text{exact}}$ in Tables 5.4 and 5.2 are very close for all but one problem. We conclude from Tables 5.3 and 5.4 that the application of adaptive cross approximations does not on average affect the quality of the computed approximations of $\mathbf{x}_{\text{exact}}$ in a significant way.

6. Conclusion. This paper describes how adaptive cross approximation can be applied in conjunction with regularization methods for the solution of ill-posed problems. Special variants for symmetric problems are discussed. Computed examples illustrate the feasibility of the use of adaptive cross approximation for this application.

TABLE 5.2
Errors in regularized solution with ACA and Tikhonov regularization.

Example	n	ε	k	$\ A - M_k\ $	$\ \mathbf{x}_k^{(\mu)} - \mathbf{x}\ /\ \mathbf{x}\ $	μ
Ex. 5.1	2 000	1 e−04	6	1.59 e−06	1.38 e−01	4.05 e−05
Ex. 5.2	2 000	1 e−04	12	1.39 e−05	4.71 e−02	4.10 e−05
Ex. 5.2	14 500	1 e−04	12	1.40 e−05	4.70 e−02	4.04 e−05
Ex. 5.2s	2 000	1 e−04	12	1.13 e−05	4.70 e−02	4.14 e−05
Ex. 5.3	2 000	1 e−02	6	9.76 e−05	2.36 e−02	5.47 e−05
Ex. 5.3s	2 000	1 e−02	6	5.58 e−03	8.23 e−02	4.44 e−05
Ex. 5.4	2 000	1 e−02	14	3.36 e−02	4.81 e−02	6.10 e−05
Ex. 5.4s	2 000	1 e−02	13	3.93 e−03	9.04 e−02	6.02 e−05

TABLE 5.3
Errors in regularized solution with TSVD regularization without ACA.

Example	n	ε	$\ \mathbf{x}_k - \mathbf{x}\ /\ \mathbf{x}\ $	k
Ex. 5.1	2 000	1 e−04	6.04 e−02	5
Ex. 5.2	2 000	1 e−04	3.22 e−02	9
Ex. 5.3	2 000	1 e−02	3.12 e−02	2
Ex. 5.4	2 000	1 e−02	2.09 e−02	8

In particular, the examples show that the quality of the computed approximations of the desired solution $\mathbf{x}_{\text{exact}}$ is not adversely affected by the reduction of the original large problem to a smaller one by adaptive cross approximation.

REFERENCES

- [1] M. L. Baart, *The use of auto-correlation for pseudo-rank determination in noisy ill-conditioned least-squares problems*, IMA J. Numer. Anal., 2 (1982), pp. 241–247.
- [2] J. Baglama and L. Reichel, *An implicitly restarted block Lanczos bidiagonalization method using Leja shifts*, BIT Numer. Math., 53 (2013), pp. 285–310.
- [3] M. Bebendorf, *Approximation of boundary element matrices*, Numer. Math., 86 (2000), pp. 565–589.
- [4] M. Bebendorf and R. Grzibovski, *Accelerating Galerkin BEM for linear elasticity using adaptive cross approximation*, Math. Methods Appl. Sci., 29 (2006), pp. 1721–1747.
- [5] M. Bebendorf and S. Rjasanow, *Adaptive low-rank approximation of collocation matrices*, Computing, 70 (2003), pp. 1–24.
- [6] C. Brezinski, G. Rodriguez, and S. Seatzu, *Error estimates for linear systems with applications to regularization*, Numer. Algorithms, 49 (2008), pp. 85–104.
- [7] D. Calvetti, J. Petersen, and L. Reichel, *A parallel implementation of the GMRES algorithm*, in Numerical Linear Algebra, eds. L. Reichel, A. Ruttan, and R. S. Varga, de Gruyter, Berlin, 1993, pp. 31–46.
- [8] D. Calvetti and L. Reichel, *Tikhonov regularization of large linear problems*, BIT Numer. Math., 43 (2003), pp. 263–283.
- [9] J. W. Daniel, W. B. Gragg, L. Kaufman, and G. W. Stewart, *Reorthogonalization and stable algorithms for updating the Gram–Schmidt QR factorization*, Math. Comp., 30 (1976), pp. 772–795.
- [10] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.
- [11] J. Erhel, *A parallel GMRES version for general sparse matrices*, Electron. Trans. Numer. Anal., 3 (1995), pp. 160–176.
- [12] L. Fox and E. T. Goodwin, *The numerical solution of non-singular linear integral equations*,

TABLE 5.4
Errors in regularized solution with Tikhonov regularization without ACA.

Example	n	ε	$\ \mathbf{x}^{(\mu)} - \mathbf{x}\ /\ \mathbf{x}\ $	μ
Ex. 5.1	2 000	1 e−04	1.38 e−01	4.05 e−05
Ex. 5.2	2 000	1 e−04	4.70 e−02	4.12 e−05
Ex. 5.3	2 000	1 e−02	2.32 e−02	5.47 e−05
Ex. 5.4	2 000	1 e−02	8.39 e−02	6.01 e−05

- Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 245:902 (1953), pp. 501–534.
- [13] K. Frederix and M. Van Barel, *Solving a large dense linear system by adaptive cross approximation*, J. Comput. Appl. Math., 234 (2010), pp. 3181–3195.
 - [14] S. Gazzola, P. Novati, and M. R. Russo, *On Krylov projection methods and Tikhonov regularization*, Electron. Trans. Numer. Anal., 44 (2015), pp. 83–123.
 - [15] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed., Johns Hopkins University Press, Baltimore, 2013.
 - [16] S. A. Goreinov, E. E. Tyrtyshnikov, and N. L. Zamarashkin, *A theory of pseudo-skeleton approximation*, Linear Algebra Appl., 261 (1997), pp. 1–21.
 - [17] S. A. Goreinov, N. L. Zamarashkin, and E. E. Tyrtyshnikov, *Pseudo-skeleton approximations by matrices of maximal volume*, Math. Notes, 62 (1997), pp. 515–519.
 - [18] C. W. Groetsch, *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*, Pitman, Boston, 1984.
 - [19] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia, 1998.
 - [20] P. C. Hansen, *Deconvolution and regularization with Toeplitz matrices*, Numer. Algorithms, 29 (2002), pp. 323–378.
 - [21] P. C. Hansen, *Regularization Tools version 4.0 for Matlab 7.3*, Numer. Algorithms, 46 (2007), pp. 189–194. Software is available in Netlib at <http://www.netlib.org>.
 - [22] M. E. Hochstenbach, *Harmonic and refined extraction methods for the singular value problem, with applications in least-squares problems*, BIT Numer. Math., 44 (2004), pp. 721–754.
 - [23] M. E. Hochstenbach, N. McNinch, and L. Reichel, *Discrete ill-posed least-squares problems with a solution norm constraint*, Linear Algebra Appl., 436 (2012), pp. 3801–3818.
 - [24] Z. Jia and D. Niu, *An implicitly restarted refined bidiagonalization Lanczos method for computing a partial singular value decomposition*, SIAM J. Matrix Anal. Appl., 25 (2003), pp. 246–265.
 - [25] S. Kindermann, *Convergence analysis of minimization-based noise level-free parameter choice rules for linear ill-posed problems*, Electron. Trans. Numer. Anal., 38 (2011), pp. 233–257.
 - [26] S. Noschese and L. Reichel, *A modified TSVD method for discrete ill-posed problems*, Numer. Linear Algebra Appl., 21 (2014), pp. 813–822.
 - [27] L. Reichel and G. Rodriguez, *Old and new parameter choice rules for discrete ill-posed problems*, Numer. Algorithms, 63 (2013), pp. 65–87.
 - [28] C. B. Shaw, Jr., *Improvements of the resolution of an instrument by numerical solution of an integral equation*, J. Math. Anal. Appl., 37 (1972), pp. 83–112.
 - [29] E. E. Tyrtyshnikov, *Incomplete cross approximation in the mosaic-skeleton method*, Computing, 64 (2000), pp. 367–380.
 - [30] Y. Yamamoto, Y. Nakatsukasa, Y. Yanagisawa, and T. Fukaya, *Roundoff error analysis of the CholeskyQR2 algorithm*, Electron. Trans. Numer. Anal., 44 (2015), pp. 306–326.